

# Errors bounds for finite approximations of coherent lower previsions on finite probability spaces

Damjan Škulj

University of Ljubljana, Faculty of Social Sciences  
Kardeljeva pl. 5, SI-1000 Ljubljana, Slovenia  
`damjan.skulj@fdv.uni-lj.si`

September 20, 2016

## Abstract

Coherent lower previsions are general models of uncertainty in probability distributions. They are often approximated by some less general models, such as coherent lower probabilities or in terms of some finite set of constraints. The amount of error induced by the approximations has been often neglected in the literature, despite the fact that it can be quite substantial. The aim of this paper is to provide a general method for estimating the exact degree of error for given approximations of coherent lower previsions on finite probability spaces. The information on the maximal error is especially useful in cases where the approximations require a lot of effort to be calculated. Our method is based on convex analysis on the corresponding credal sets, which can be represented as convex polyhedra. It provides the exact maximal possible amount of error for a given finite approximation of a coherent lower prevision. An algorithm based on quadratic programming is also provided.

**Keywords.** lower prevision, partially specified lower prevision, credal set, convex polyhedron, quadratic programming

2010 Mathematics Subject Classification: 90C90

## 1 Introduction

Models of *imprecise probabilities* have been developed to cope with uncertainty in probability distributions. Single precise models are thus replaced by models compatible with multiple (precise) probability distributions. The advantage of such models compared to the classical precise models is that they can incorporate a lower or higher degree of uncertainty, which is represented through larger or smaller sets of compatible probability distributions.

Thus imprecise models subsume precise models as one extreme as well as the models of complete uncertainty on the other side. A review of models and applications of imprecise probabilities can be found in [2].

One of the most popular and also most general models of imprecise probabilities are *coherent lower previsions* (see e.g. [10, 14]). A coherent lower prevision  $\underline{P}$ , in general given on a measurable space  $(\mathcal{X}, \mathcal{A})$ , is an imprecise probability model based on judgements about the lower or upper expectations of a set of random variables  $\mathcal{K}$ , also called *gambles*. The judgement  $\underline{P}(f) = a$  implies that every precise probability distribution  $P$  compatible with  $\underline{P}$  must satisfy  $E_P(f) \geq a$ , that is  $\underline{P}(f)$  means that the expectation of  $f$  is at least  $a$ . *Coherence* in this context means that the judgements on the set of gambles allow, for every gamble  $f$ , the existence of at least one precise probability distribution  $P$  compatible with  $\underline{P}$  for which  $E_P(f) = \underline{P}(f)$ .

A coherent lower prevision  $\underline{P}$  specified on a set of gambles  $\mathcal{K}$  can have multiple possible extensions to a larger set, say  $\mathcal{H} \supset \mathcal{K}$ . In other words, there can be multiple coherent lower previsions that coincide on a set of gambles. In particular, a coherent lower prevision may be approximated by a more specific model, such as *coherent lower probability* (see e.g. [1]), in which case its restriction to *indicator gambles* is only known, i.e. an *indicator gamble*  $1_A$  is a map  $\mathcal{X} \rightarrow \mathbb{R}$  such that  $1_A(x)$  equals 1 if  $x \in A$  and 0 otherwise. There are variety of reasons for approximating coherent lower probabilities with less general methods. One reason is that in general there is no nice or elegant way to represent a general coherent lower prevision. Lower and upper probabilities can be much more elegant and intuitive as approximations. We must also keep in mind that coherent lower previsions, even on finite spaces, in general cannot be represented in terms of a single function or any other reasonably sized collection of values. In best case they can be represented as sets of extreme points of their credal sets, which in most cases are very large. Instead of calculating all extreme points it is sometimes more convenient to approximate a coherent lower prevision with its values on a suitable set of gambles and apply the *natural extension* for further calculations. In many cases, computations with coherent lower previsions are computationally demanding (consider for instance calculating lower prevision corresponding to imprecise Markov chains [5, 15, 16]), which makes it reasonable to keep the set  $\mathcal{K}$  of moderate size.

In this paper we analyze the following problem. Let  $\underline{P}$  be a lower prevision on a finite sample space  $\mathcal{X}$ . Its full description would in general require detailed information of its credal set, whose set of extreme points can be very large. Suppose that instead we know the values of  $\underline{P}$  on a set of gambles  $\mathcal{K}$ . The restriction  $\underline{P}|_{\mathcal{K}}$  approximates  $\underline{P}$  and the natural question arises, how good is this approximation. Given the restriction,  $\underline{P}$  is its extension, but there might be other extensions too. Therefore, we would like to know by how much can another extension deviate from  $\underline{P}$ . In other words, we want to find the maximal distance between two arbitrary extensions of a coherent

lower probability to the set of all gambles.

In our analysis of the maximal possible distance we first show that the maximal possible distance is always reached when one of the extensions is the *natural extension*. Consequently, much of the analysis is done on the credal set of the natural extension with the special emphasis on its extreme points. Our main theoretical result gives the exact upper bound for the maximal distance in terms of distances between the extreme points. We also provide an algorithm for finding the maximal possible distance.

The paper is structured as follows. In Section 2 we review basic concepts of imprecise probabilities with the emphasis on coherent lower previsions. In Section 3 we analyse basic properties of credal sets as convex polyhedra and apply some general concepts of convex analysis to the case of credal sets. Our main theoretical results are in Section 4. The algorithm for calculating the maximal possible distance is described in Section 5. The paper concludes with Section 6.

## 2 Notation and basic results

In this section we introduce the notation and review the concepts used in the paper. When possible we will stick with the standard terminology used in the theory of imprecise probabilities, which will sometimes be supplemented by the standard terminology of convex analysis, linear algebra and optimization.

The object of our analysis will be *coherent lower previsions* which are one of the most general models used to represent partially specified probabilities. They encompass several particular models, such as *coherent lower and upper probabilities*, *2- and n-monotone capacities*, *belief and plausibility functions*, *lower expectation functionals* and others. Mathematically, coherent lower previsions are superlinear functionals that can be equivalently represented as lower envelopes of expectation functionals.

*Gambles*. Throughout this paper let  $\mathcal{X}$  represent a finite set, a *sample space*, and  $\mathcal{L}$  the set of all real-valued maps on  $\mathcal{X}$ , also called *gambles*. Equivalently,  $\mathcal{L}$  may be viewed as the set of vectors in  $\mathbb{R}^{|\mathcal{X}|}$ . By  $1_A$  we will denote the *indicator gamble* of a set  $A \subseteq \mathcal{X}$ :

$$1_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

We will write  $1_x$  instead of  $1_{\{x\}}$  for elements  $x \in \mathcal{X}$ .

The set of gambles will be endowed by the standard inner product

$$f \cdot g = \sum_{x \in \mathcal{X}} f(x)g(x), \quad (2)$$

which generates the  $l^2$  norm:

$$\|f\| = \sqrt{f \cdot f} = \sqrt{\sum_{x \in \mathcal{X}} f(x)^2}, \quad (3)$$

and the Euclidean distance between vectors:

$$d(f, g) = \|f - g\|, \quad (4)$$

which will be used by default throughout the paper.

*Linear previsions.* A *linear prevision*  $P$  is an expectation functional with respect to some probability mass vector  $p$  on  $\mathcal{X}$ . It maps a gamble  $f$  into a real number  $P(f)$ . Usually, we will write

$$P(f) = \sum_{x \in \mathcal{X}} p(x)f(x) =: P \cdot f. \quad (5)$$

The set of linear previsions is therefore a subset of the dual space of  $\mathcal{L}$ .

The inner product notation used on the right hand side of the above equation is introduced because we will often use linear functionals of the form  $f \mapsto p \cdot f$  where the vector  $p$  will not necessarily be a probability mass vector. We will then use the inner product notation to avoid misinterpretations. Without danger of confusion we will therefore interpret a linear prevision  $P$  as a vector with the same length as gambles in  $\mathcal{L}$ .

*Probability simplex.* If the sample space  $\mathcal{X}$  contains exactly three elements, say  $\mathcal{X} = \{x, y, z\}$ , the probability mass vectors can be represented as points of the form  $(p(x), p(y), p(z))$  in  $\mathbb{R}^3$ . However, since the restriction  $p(x) + p(y) + p(z)$  applies, they in fact form a two dimensional space, which can be depicted as an equilateral triangle with vertices  $x, y$  and  $z$ . Given any point in this triangle, the sum of distances to its sides is constantly equal to its altitude, which equals  $\frac{\sqrt{3}}{2}a$ , where  $a$  is the common length of the sides. Taking  $a = \frac{2}{\sqrt{3}}$  makes the altitude equal to 1. The distance of a point from each side now denotes the probability of the point in the opposite vertex (see Fig. 1). Probability simplex diagrams are very useful to illustrate concepts of imprecise probabilities; however, one needs to be cautious not to be misled by specifics of low dimensional probability spaces.

*Coherent lower previsions.* A *coherent lower prevision* on an arbitrary set of gambles  $\mathcal{K}$  is a mapping  $\underline{P}: \mathcal{K} \rightarrow \mathbb{R}$  that allows the representation

$$\underline{P}(f) = \min_{P \in \mathcal{M}(\underline{P})} P(f) \quad (6)$$

for every  $f \in \mathcal{K}$ , where  $\mathcal{M}(\underline{P})$  is a closed and convex set of linear previsions. The set  $\mathcal{M}(\underline{P})$  is called a *credal set* of  $\underline{P}$ . We will often denote a credal set just by  $\mathcal{M}$ .

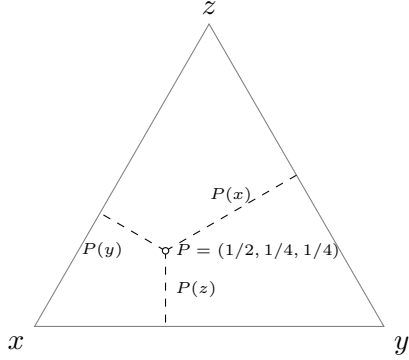


Figure 1: Probability simplex: the distance from a side denotes the probability of the element at the opposite vertex.

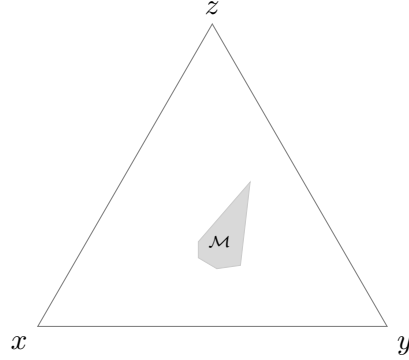


Figure 2: The credal set  $\mathcal{M}$  of the lower prevision from Example 1.

*The natural extension.* Given a coherent lower prevision  $\underline{P}$  on  $\mathcal{K}$ , it is possible to extend it to the set of all gambles  $\mathcal{L}$  in possibly several different ways. However, there is a unique minimal extension, called the *natural extension*:

$$\underline{E}(f) = \min_{P \in \mathcal{M}(\underline{P})} P(f). \quad (7)$$

As the natural extension is the *lower envelope* or the *support function* of a credal set, containing expectation functionals, we may call a coherent lower prevision defined on the entire  $\mathcal{L}$  a *lower expectation functional*.

A mapping  $\underline{P}: \mathcal{K} \rightarrow \mathbb{R}$  where  $\mathcal{K}$  is a linear (vector) space is a coherent lower prevision if and only if it satisfies the following axioms ([10]) for all  $f, g \in \mathcal{K}$  and  $\lambda \geq 0$ :

(P1)  $\underline{P}(f) \geq \inf_{x \in \mathcal{X}} f(x)$  [accepting sure gains];

(P2)  $\underline{P}(\lambda f) = \lambda \underline{P}(f)$  [positive homogeneity];

(P3)  $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$  [superlinearity].

An easy consequence of the definitions is *constant additivity*:

$$\underline{P}(f + \lambda 1_{\mathcal{X}}) = \underline{P}(f) + \lambda \quad (8)$$

for any  $\lambda \in \mathbb{R}$ .

### 3 Convex analysis on credal sets

#### 3.1 Credal set as a closed convex set

A credal set is a closed and convex set of linear previsions. Since every linear prevision can be uniquely represented as a probability mass vector, a credal

set can be represented as a convex set of probability mass vectors. The set  $\mathcal{M}$  is therefore the maximal set of  $|\mathcal{X}|$ -dimensional vectors  $p$  satisfying

$$p \cdot f \geq \underline{P}(f) \quad \text{for every } f \in \mathcal{K}, \quad (9)$$

$$p \cdot 1_x \geq 0 \quad \text{for every } x \in \mathcal{X} \text{ and} \quad (10)$$

$$p \cdot 1_{\mathcal{X}} = 1. \quad (11)$$

In the case where  $\mathcal{K}$  is finite,  $\mathcal{M}$  is bounded by a finite number of support hyperplanes. Replacing  $f$  by  $f - \underline{P}(f)$ , we can achieve that  $\underline{P}(f - \underline{P}(f)) = 0$ , which follows by constant additivity. The constraints of the form

$$p \cdot (f - \underline{P}(f)) \geq 0 \quad (12)$$

are then equivalent to (9). Thus, when needed, we may without loss of generality assume that  $\underline{P}(f) = 0$  for every  $f \in \mathcal{K}$ .

In the case where  $\mathcal{K}$  is finite, the corresponding credal set is a *convex polyhedron*. Strictly speaking, it is an  $\mathcal{H}$ -polyhedron, which means that it is bounded and an intersection of a finite number of half spaces. According to Theorem 14.3 in [8] every  $\mathcal{H}$ -polyhedron in an  $\mathbb{R}^m$  is also a  $\mathcal{V}$ -polyhedron, which means that it is a convex combination of a finite number of extreme points. The properties of credal sets as convex polyhedra have also been recently studied by Quaeghebeur [13].

From now on we will call credal sets that are convex polyhedra *finitely generated credal sets*. Similarly, we will denote coherent lower previsions defined on finite sets of gambles or their natural extensions *finitely generated coherent lower previsions*.

**Example 1.** Let  $\underline{P}$  be a lower prevision on  $\mathcal{K} = \{f_1, \dots, f_5\}$  where

$$\begin{aligned} f_1 &= (0, 1, 0.5) & f_2 &= (0, 0.5, 1) \\ f_3 &= (0.15, 0, 1) & f_4 &= (1, 0, 0.6) \\ f_5 &= (0.2, 1, 0) \end{aligned}$$

and

$$\begin{aligned} \underline{P}(f_1) &= 0.46 & \underline{P}(f_2) &= 0.4 & \underline{P}(f_3) &= 0.25 \\ \underline{P}(f_4) &= 0.44 & \underline{P}(f_5) &= 0.4 \end{aligned}$$

The credal set corresponding to  $\underline{P}$  is depicted in Figure 2.

*Faces and extreme points of a finitely generated credal set.* Let  $\underline{E}$  be a natural extension of a coherent lower prevision on a finite set  $\mathcal{K}$ . The corresponding credal set  $\mathcal{M}$  is then finitely generated by constraints (9)–(11). Except for (11), all these constraints can be written in the form  $p \cdot f \geq$

0, some of which may also be *loose*, which means that equality cannot be reached for any point of the credal set. Since coherence is assumed, which means exactly that equality  $P(f) = \underline{P}(f)$  can be reached for every  $f \in \mathcal{K}$  and some  $P \in \mathcal{M}$ , the loose constraints can only be found among (10). After removing all loose constraints, we may represent the set  $\mathcal{M}$  as the set of all  $|\mathcal{X}|$ -dimensional vectors  $p$  satisfying the constraints

$$p \cdot f_i \geq 0 \text{ for } i = 1, \dots, n, \quad (13)$$

$$p \cdot 1_{\mathcal{X}} = 1. \quad (14)$$

The set  $\{f_1, \dots, f_n\}$  thus contains all elements of  $\mathcal{K}$  together with possibly some indicator gambles of the form  $1_x$ . Moreover, we will from now on assume that none of the above constraints are loose. We will denote  $\mathcal{K}^+ = \{f_1, \dots, f_n\}$ . In most cases, however, we will assume that  $\mathcal{K}^+ = \mathcal{K}$ .

The *faces* of a credal set  $\mathcal{M}$  are the sets of the form

$$\mathcal{M}_f = \{P \in \mathcal{M} : P(f) = \underline{P}(f)\}, \quad (15)$$

where  $f$  is an arbitrary gamble. The smallest faces are exactly the extreme points and the faces of codimension 1 are called *facets*<sup>1</sup>. The set of all extreme points of  $\mathcal{M}$  will be denoted by  $\mathcal{E}(\mathcal{M})$  or simply  $\mathcal{E}$ . The set of extreme points of a face  $\mathcal{M}_f$  will be denoted by  $\mathcal{E}_f$ , and  $\mathcal{E}_f \subseteq \mathcal{E}$  holds.

**Example 2.** The extreme points of the credal set from Example 1 are

$$\begin{aligned} E_1 &= (0.4, 0.32, 0.28) & E_2 &= (0.43, 0.35, 0.23) \\ E_3 &= (0.39, 0.42, 0.19) & E_4 &= (0.32, 0.48, 0.20) \\ E_5 &= (0.15, 0.37, 0.48) \end{aligned}$$

(See Figure 4.)

We extend a credal set  $\mathcal{M}$  to the set of vectors

$$\hat{\mathcal{M}} = \{p : p \cdot f \geq 0, \text{ for every } f \in \mathcal{K}^+\}, \quad (16)$$

which is a convex cone, with the *basis*  $\mathcal{M}$ . This means that every  $p \in \hat{\mathcal{M}}$  is of the form  $p = \lambda P$  for some  $\lambda \geq 0$  and  $P \in \mathcal{M}$ .

Given a credal set  $\mathcal{M}$ , the *cone of desirable gambles* contains exactly those gambles in  $\mathcal{L}$  whose lower prevision is non-negative:

$$\mathcal{D} = \{f \in \mathcal{L} : p \cdot f \geq 0 \text{ for every } p \in \mathcal{M}\}. \quad (17)$$

The gambles  $f$  with  $\underline{P}(f) = 0$  are sometimes called *marginally desirable*.

---

<sup>1</sup>The codimension 1 is meant relative to the dimension of  $\mathcal{M}$ . That is  $\dim \mathcal{M}_f = \dim \mathcal{M} - 1$ . Note also that a credal set is at most of dimension  $|\mathcal{X}| - 1$  because of the constraint  $P(1_{\mathcal{X}}) = 1$ .

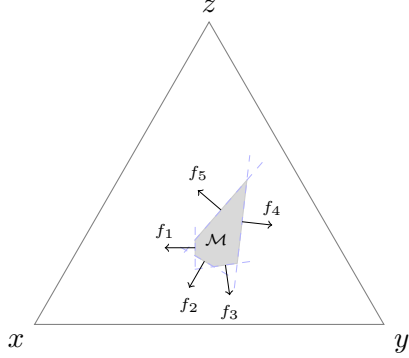


Figure 3: Credal set from Example 1 as an intersection of half planes: their support lines are dashed, gambles  $f_i \in \mathcal{K}^+$  are depicted as normal vectors to faces

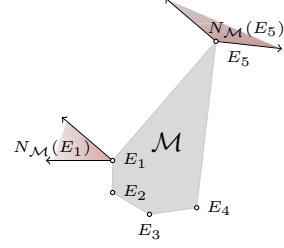


Figure 4: Normal cones at extreme points are the positive hulls of the normal vectors of adjacent faces.

### 3.2 Normal cones of credal sets

From now on we will assume that the set of gambles  $\mathcal{K}$  together with the corresponding restriction of the lower prevision contains exactly the gambles required to represent the credal set corresponding to its natural extension in the form of (13). In this section we will provide some results that will allow us to calculate the maximal normed distance  $\frac{|P(h)-E(h)|}{\|h\|}$ , conditional on  $E(h) = \underline{P}(h)$ , where  $E$  and  $P$  are extreme points of the credal set  $\mathcal{M}(\underline{P})$ .

*The normal cone.* Let

$$\mathcal{C} = \{p: p \cdot f \geq \underline{P}(f), \forall f \in \mathcal{K}\} \quad (18)$$

be a convex polyhedron, where  $\mathcal{K}$  is a set of vectors and  $\underline{P}$  a real valued map on  $\mathcal{K}$ , and let  $p$  be one of its boundary points. The *normal cone* of  $\mathcal{C}$  at  $p$  is the set

$$N_{\mathcal{C}}(p) = \{f: p \cdot f \geq q \cdot f \text{ for every } q \in \mathcal{C}\}. \quad (19)$$

The normal cone of a credal set at some point  $P \in \mathcal{M}$  is thus the set of gambles  $f$  that satisfy  $P(f) = \underline{P}(f)$ .

**Proposition 1** ([8] Proposition 14.1.). *Let  $\mathcal{C}$  be a convex polyhedron of the form (18) and  $p \in \mathcal{C}$  an element on its boundary. Let  $p \cdot f_i = \underline{P}(f_i)$  hold for exactly those  $f_i \in \mathcal{K}$ , where  $i \in I \subseteq \{1, \dots, n\}$ . Then*

$$N_{\mathcal{C}}(p) = \text{pos} \{f_i: i \in I\},$$

where *pos* denotes the positive hull.

**Remark 1.** We will call the set of vectors  $\{f_i: i \in I\}$  the *positive basis* of the normal cone  $N_{\mathcal{C}}(p)$ .



**Corollary 1.** *Let  $\mathcal{M}$  be a credal set defined by constraints (13) and (14). Then the set of desirable gambles  $\mathcal{D}$  corresponding to  $\mathcal{M}$  is the normal cone of  $\hat{\mathcal{M}}$  at  $\mathbf{0}$  and we have that*

$$\mathcal{D} = \text{pos} \{f : f \in \mathcal{K}\}. \quad (20)$$

*Proof.* The set  $\hat{\mathcal{M}}$  is a convex cone whose support hyperplanes are exactly the sets of the form  $H_f = \{p : p \cdot f = 0\}$  and the origin is exactly the intersection of all support hyperplanes:  $\mathbf{0} \cdot f = 0$  for every  $f \in \mathcal{K}^+$ . We can therefore apply Proposition 1.  $\square$

**Remark 2.** In [2] Chapter 1, the set constructed as in (20) is called the natural extension of the assessment  $\mathcal{K}$ . The fact that the set of desirable gambles is the positive hull of marginally desirable assessments in  $\mathcal{K}$  with included strictly positive gambles can also be found in Chapter 2 of the mentioned book.

**Corollary 2.** *Let  $\mathcal{M}$  be a credal set defined by constraints (13) and (14),  $E \in \mathcal{M}$  a linear prevision and  $h$  a gamble such that  $E(h) = \underline{P}(h)$ . Suppose that  $E(f_i) = 0$  for exactly  $i \in I \subseteq \{1, \dots, n\}$ . Then there exist  $\alpha_i \geq 0$  for every  $i \in I$  and  $\beta \in \mathbb{R}$  so that*

$$h = \sum_{i \in I} \alpha_i f_i + \beta 1_{\mathcal{X}}. \quad (21)$$

*Proof.* Let  $h \in \mathcal{L}$  be a gamble such that  $E(h) = \underline{P}(h)$ . Set  $g = h - \underline{P}(h)$ . Then, for every  $p \in \hat{\mathcal{M}}$  (see (16)),  $p = \alpha P$  for some  $P \in \mathcal{M}$  and  $\alpha \geq 0$ . Therefore  $p \cdot g = \alpha P \cdot g \geq 0 = E \cdot g$ , whence  $g \in N_{\hat{\mathcal{M}}}(E)$ . By Proposition 1,  $g = \sum_{i \in I} \alpha_i f_i$  for some positive constants  $\alpha_i$ . Hence  $h = \sum_{i \in I} \alpha_i f_i + \underline{P}(h) 1_{\mathcal{X}}$ , which proves the proposition.  $\square$

*Minimum norm elements of the normal cone.* Consider an element  $h$  of the form (21). Given a pair of expectation functionals  $E$  and  $P$ , the distance  $P(h) - E(h)$  does not depend on  $\beta$ . In order to maximize the normed distance, we must consider the representative with the minimum norm, as the norm appears in the denominator of the expression. The characterization of the minimal norm element of the form (21) follows.

**Proposition 2.** *Let  $h$  be a gamble. Then  $\|h + \beta 1_{\mathcal{X}}\| \geq \|h\|$  for every  $\beta \in \mathbb{R}$  if and only if  $h \cdot 1_{\mathcal{X}} = 0$ .*

*Proof.* We have that  $\|h + \beta 1_{\mathcal{X}}\|^2 = \|h\|^2 + \beta^2 + 2\beta h \cdot 1_{\mathcal{X}}$ , which has minimum in  $\beta = h \cdot 1_{\mathcal{X}}$ . Hence the minimizing  $\beta$  equals 0 exactly if  $h \cdot 1_{\mathcal{X}}$  does.  $\square$

**Corollary 3.** *Let  $E, h$  and  $I$  be as in Corollary 2 and let  $f'_i$  be the unique vectors such that  $f_i - f'_i = c 1_{\mathcal{X}}$  and  $f'_i \cdot 1_{\mathcal{X}} = 0$  for every  $i \in I$ . Then there exist some  $\alpha'_i \geq 0$  for every  $i \in I$  and  $\beta' \in \mathbb{R}$  so that*

$$h = \sum_{i \in I} \alpha'_i f'_i + \beta' 1_{\mathcal{X}}. \quad (22)$$

Moreover,

$$\left\| \sum_{i \in I} \alpha'_i f'_i \right\| \leq \left\| \sum_{i \in I} \alpha'_i f'_i + \beta 1_{\mathcal{X}} \right\| \text{ for every } \beta \in \mathbb{R}. \quad (23)$$

*Proof.* Since  $f'_i \cdot 1_{\mathcal{X}} = 0$ , we have that  $(\sum_{i \in I} \alpha'_i f'_i) \cdot 1_{\mathcal{X}} = 0$ , whence by Proposition 2 it follows that this is the minimal-norm gamble of the form (22).  $\square$

Let  $I$  and  $f'_i$ , for  $i \in I$ , be as in Corollary 2 and let  $\underline{\alpha}: I \rightarrow [0, \infty)$  be a map and  $\beta \in \mathbb{R}$  a constant (we will write  $\alpha_i$  instead of  $\alpha(i)$ ). Then we define  $h(\underline{\alpha}, \beta) = \sum_{i \in I} \alpha_i f'_i + \beta 1_{\mathcal{X}}$ . Clearly,  $h(\underline{\alpha}, \beta) \in N_{\mathcal{M}}(E)$  and every element of  $N_{\mathcal{M}}(E)$  is of the form  $h(\underline{\alpha}, \beta)$ .

**Corollary 4.** *The following equality holds:*

$$\max_{(\underline{\alpha}, \beta)} \frac{|E(h(\underline{\alpha}, \beta)) - P(h(\underline{\alpha}, \beta))|}{\|h(\underline{\alpha}, \beta)\|} = \max_{\underline{\alpha}} \frac{|E(h(\underline{\alpha}, 0)) - P(h(\underline{\alpha}, 0))|}{\|h(\underline{\alpha}, 0)\|} \quad (24)$$

*Proof.* Since  $|E(h + \beta 1_{\mathcal{X}}) - P(h + \beta 1_{\mathcal{X}})| = |E(h) - P(h)|$ , the maximum of the expression is achieved at  $h$  with the minimum norm, which is the one with  $\beta = 0$ .  $\square$

*Maximal distance between expectation functionals.* Take two linear expectation functionals  $P$  and  $E \in \mathcal{M}$  and let  $I$  and  $f'_i$  for  $i \in I$  be as in Corollary 2. Our goal is to find the normed distance

$$d_E(E, P) = \max_{h \in N_{\mathcal{M}}(E)} \frac{|P(h) - E(h)|}{\|h\|} = \max_{h \in N_{\mathcal{M}}(E)} \frac{P(h) - E(h)}{\|h\|}. \quad (25)$$

The absolute value can be omitted because  $E(h) = \min_{P \in \mathcal{M}} P(h)$  for every  $h \in N_{\mathcal{M}}(E)$ . By Corollary 4, every  $h \in N_{\mathcal{M}}(E)$  that can minimize the above expression is of the form  $h(\underline{\alpha}, 0)$ . Since  $E$  and  $P$  are themselves vectors too, we can denote  $D = P - E$ , and write

$$P(h) - E(h) = (P - E) \cdot h = D \cdot h.$$

Now we can decompose every  $f_i$  for  $i \in I$  as

$$f_i = \lambda_i D + u_i, \quad (26)$$

so that  $D \cdot u_i = 0$ . Given that  $h = \sum_{i \in I} \alpha_i f_i$ , we obtain

$$h = (\underline{\alpha} \cdot \underline{\lambda}) D + \underline{\alpha} \cdot U, \quad (27)$$

where  $U$  is the matrix whose rows are  $u_i$ ,  $\underline{\lambda}$  is the column vector with components  $\lambda_i$  and the vectors  $f'_i$  are also written as row vectors. We also assume  $\underline{\alpha}$  to be a column vector.

Further we have that

$$\|h\|^2 = \|D\|^2 \underline{\alpha} \underline{\lambda} \underline{\lambda}^t \underline{\alpha}^t + \underline{\alpha} U U^t \underline{\alpha}^t. \quad (28)$$

Now denote  $\Pi = \|D\|^2 \underline{\lambda} \underline{\lambda}^t + U U^t$  and write:

$$\|h\|^2 = \underline{\alpha} \Pi \underline{\alpha}^t. \quad (29)$$

Clearly,  $\Pi$  is a symmetric and positive semi-definite matrix.

Moreover, we have that

$$P(h) - E(h) = D \cdot (\underline{\alpha} \cdot \underline{\lambda}) D = (\underline{\alpha} \cdot \underline{\lambda}) \|D\|^2. \quad (30)$$

Our goal is the maximization of expression (25). Thus we need to maximize

$$\varphi(\underline{\alpha}) = \frac{(\underline{\alpha} \cdot \underline{\lambda}) \|D\|^2}{\sqrt{\underline{\alpha} \Pi \underline{\alpha}^t}} \quad (31)$$

over the set of all  $I$ -vectors  $\underline{\alpha}$  with non-negative components. Clearly, for every non negative constant  $k$  we have that  $\varphi(k\underline{\alpha}) = \varphi(\underline{\alpha})$ . Moreover, only those  $\underline{\alpha}$  for which the numerator in  $\varphi(\underline{\alpha})$  is positive are of interest, and then multiplying  $\underline{\alpha}$  by a suitable positive constant can ensure that the numerator is 1. Maximizing  $\varphi(\underline{\alpha})$  is then equivalent to minimizing the nominator, which yields the following quadratic programming problem:

Minimize:

$$\underline{\alpha} \Pi \underline{\alpha}^t \quad (32)$$

subject to

$$(\underline{\alpha} \cdot \underline{\lambda}) \|D\|^2 = 1 \quad (33)$$

$$\underline{\alpha} \geq 0 \quad (34)$$

**Example 3.** Consider the lower prevision  $\underline{P}$  from Example 1. We will calculate the distance  $d_{E_1}(E_1, E_5)$ , where  $E_1 = (0.4, 0.32, 0.28)$  and  $E_5 = (0.15, 0.37, 0.48)$ . First we have:

$$D = E_5 - E_1 = (-0.2462, 0.0492, 0.1969),$$

and its norm, which is the Euclidean distance between the two extreme points is  $\|D\| = 0.3191$ . The positive basis of  $N_{\mathcal{M}}(E_1)$  consists of the transformed gambles

$$\begin{aligned} f'_1 &= f_1 - f_1 \cdot 1_{\mathcal{X}}/3 = (-0.5, 0.5, 0) \\ f'_5 &= f_5 - f_5 \cdot 1_{\mathcal{X}}/3 = (-0.2, 0.6, -0.4). \end{aligned}$$

(see Corollary 3).

We have  $f'_1 = 1.451D + (-0.1429, 0.4286, -0.2857)$ , and since  $f'_5$  is orthogonal to  $D$ , it follows that  $u_5 = f'_5$  and  $\lambda_2 = 0$ . Thus

$$\underline{\lambda} = \begin{bmatrix} 1.451 \\ 0 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} -0.14 & 0.43 & -0.29 \\ -0.20 & 0.60 & -0.40 \end{bmatrix}$$

which gives

$$\Pi = \|D\|^2 \underline{\lambda} \underline{\lambda}^t + U U^t = \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.56 \end{bmatrix}$$

Taking  $\underline{\alpha} = (\alpha_1, \alpha_2)^t$ , we obtain the objective function to be minimized:

$$\underline{\alpha} \Pi \underline{\alpha}^t = 0.5\alpha_1^2 + 0.8\alpha_1\alpha_2 + 0.56\alpha_2^2$$

subject to

$$\|D\|^2 \underline{\alpha} \cdot \underline{\lambda} = \|D\|^2 \lambda_1 \alpha_1 = 1$$

whence  $\alpha_1 = 6.7708$ . Substituting  $\alpha_1$  in the objective function we obtain

$$\underline{\alpha} \Pi \underline{\alpha}^t = 22.9219 + 5.41664\alpha_2 + 0.56\alpha_2^2,$$

which has to be minimized subject to  $\alpha_2 \geq 0$ . The minimum is obtained for  $\alpha_2 = 0$ , with the minimal value of objective function  $\underline{\alpha} \Pi \underline{\alpha}^t$  equal to 22.9219. Now

$$d_{E_1}(E_1, E_5) = \varphi(\underline{\alpha}) = 1/\sqrt{22.9219} = 0.2089.$$

Note that this is significantly less than the Euclidean distance between the points, which is equal to  $\|D\| = 0.3191$ .

The results in the example show that the maximal normed distance is reached on the gamble that makes the smallest angle with  $D$  among all gambles in the normal cone, and in our case this is obviously  $f'_1$ , since adding any positive part of  $f'_5$ , which is perpendicular to  $D$ , would only increase the angle. The normed distance is in general clearly bounded by the Euclidean distance, which is reached only in the case where the normal cone contains a vector that is parallel with  $D$ . In most cases, however, particularly when the normal cones are narrow, the normed distance is typically significantly lower than the Euclidean distance.

## 4 Maximal distance between extensions of finitely generated coherent lower previsions

We are now going to use the results developed in the previous section to estimate the maximal positive normed distance between two lower previsions coinciding on a finite set  $\mathcal{K}$ . As we will see, we can assume one of them to be the natural extension of the restriction to  $\mathcal{K}$ . In Figure 5 two credal sets are depicted in the probability simplex that coincide on a set of gambles.

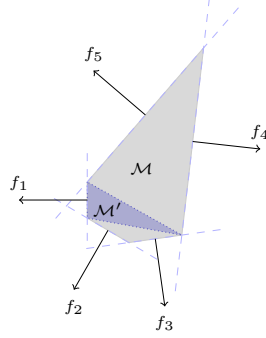


Figure 5: Lower previsions  $\underline{P}$  and  $\underline{P}'$  with the credal sets  $\mathcal{M}$  and  $\mathcal{M}'$  respectively coincide on the set of gambles  $\mathcal{K} = \{f_1, \dots, f_5\}$ . (Note that  $\underline{P}$  is the natural extension of  $\underline{P}|_{\mathcal{K}}$ .)

#### 4.1 The distance between coherent lower previsions

Let  $\underline{P}$  and  $\underline{P}'$  be two coherent lower previsions on the set of all gambles  $\mathcal{L}$  on a finite set  $\mathcal{X}$ . We define the distance<sup>2</sup> between  $\underline{P}$  and  $\underline{P}'$  as

$$d(\underline{P}, \underline{P}') = \max_{f \in \mathcal{L}} \frac{|\underline{P}(f) - \underline{P}'(f)|}{\|f\|}, \quad (35)$$

where the norm  $\|f\| = \sqrt{f \cdot f}$  is the Euclidian norm in  $\mathbb{R}^{|\mathcal{X}|}$ . Clearly, the following alternative definition is equivalent:

$$d(\underline{P}, \underline{P}') = \max_{\substack{f \in \mathcal{L} \\ \|f\|=1}} |\underline{P}(f) - \underline{P}'(f)|, \quad (36)$$

It is readily verified that the above distance function induces a metric in the set of all lower previsions on  $\mathcal{L}$ . In this section we will analyse the maximal possible distance between two coherent lower previsions that coincide on a finite set of gambles.

Suppose that  $\underline{P}$  is a lower prevision on  $\mathcal{L}$ , and the only information about it are the values on a finite set of gambles  $\mathcal{K} \subset \mathcal{L}$ . That is  $\underline{P}(f)$  for every  $f \in \mathcal{K}$  are given. As argued above, it is convenient to assume that  $\underline{P}(f) = 0$  for every  $f \in \mathcal{K}$ . The natural extension  $\underline{E}$  is the minimal (or the least committal) extension of  $\underline{P}_{\mathcal{K}}$ . This implies that  $\underline{P}(f) \geq \underline{E}(f)$  for every  $f \in \mathcal{L}$ . Therefore, given another extension  $\underline{P}'$  of  $\underline{P}_{\mathcal{K}}$ , we have that

$$|\underline{P}(f) - \underline{P}'(f)| \leq \max\{\underline{P}(f) - \underline{E}(f), \underline{P}'(f) - \underline{E}(f)\}, \quad (37)$$

which implies that

$$d(\underline{P}, \underline{P}') \leq \max\{d(\underline{P}, \underline{E}), d(\underline{P}', \underline{E})\}. \quad (38)$$

<sup>2</sup>For another distance function between coherent lower previsions, see e.g. [17].

As we are interested in the maximal possible distance between coherent lower previsions coinciding on  $\mathcal{K}$ , it will therefore be enough to focus to the case where one of them is the natural extension of  $\underline{P}_{\mathcal{K}}$ .

## 4.2 Maximal distance to the natural extension

Let  $\underline{E}$  and  $\underline{P}$  be respectively the natural extension of  $\underline{P}_{\mathcal{K}}$  and another extension, and  $\mathcal{M}$  and  $\mathcal{C}$  respectively their credal sets. As described in previous sections, both credal sets are convex polyhedra with extreme points  $\mathcal{E}(\mathcal{M})$  and  $\mathcal{E}(\mathcal{C})$  respectively.

Assuming the above notations, we start with the following proposition.

**Proposition 3.** *Take some  $f \in \mathcal{K}$  and let  $\mathcal{M}_f$  be the corresponding face of  $\mathcal{M}$ . Then  $\mathcal{C} \cap \mathcal{M}_f \neq \emptyset$ .*

*Proof.* Clearly,  $\mathcal{M}_f$  contains exactly all linear previsions  $P$  in  $\mathcal{M}$  such that  $P(f) = \underline{P}(f)$ . If no  $P \in \mathcal{C}$  belongs to  $\mathcal{M}_f$ , this then implies that  $P(f) > \underline{P}(f)$  for every  $P \in \mathcal{C}$ , and since  $\mathcal{C}$  is compact, this would imply that  $\min_{P \in \mathcal{C}} P(f) > \underline{P}(f)$ , which contradicts the assumptions.  $\square$

**Corollary 5.** *Let  $h \in \mathcal{L}$  be an arbitrary gamble. Then:*

- (i)  $\underline{P}(h) \leq \max_{P \in \mathcal{M}_f} P(h)$  for every  $f \in \mathcal{K}$ ;
- (ii)  $\underline{P}(h) \leq \min_{f \in \mathcal{K}} \max_{P \in \mathcal{M}_f} P(h)$ ; the inequality is tight in the sense that for every  $h \in \mathcal{L}$  an extension of  $\underline{P}_{\mathcal{K}}$  exists that gives equality in the equation.
- (iii)  $\underline{P}(h) \leq \min_{f \in \mathcal{K}} \max_{E \in \mathcal{E}_f} E(h)$  where  $\mathcal{E}_f$  is the set of extreme points of the face  $\mathcal{M}_f$ ; and the inequality is again tight.

*Proof.* (i) is an immediate consequence of Proposition 3.

The inequality in (ii) is a direct consequence of (i). It remains to prove that there is an extension of  $\underline{P}_{\mathcal{K}}$  where the equality is reached.

Let  $\mathcal{M}_f$  be a face of  $\mathcal{M}$  and let  $P_f = \arg \max_{P \in \mathcal{M}_f} P(h)$ . Let  $\mathcal{M}'$  be the convex hull of  $\{P_f : f \in \mathcal{K}\}$  and  $\underline{P}'$  the corresponding coherent lower prevision, which coincides with  $\underline{P}$  on  $\mathcal{K}$  by construction. For every  $P \in \mathcal{M}'$  we have that  $P = \sum_{f \in \mathcal{K}} \alpha_f P_f$ , for some collection of values  $\alpha_f \geq 0$  for every  $f \in \mathcal{K}$  and  $\sum_{f \in \mathcal{K}} \alpha_f = 1$ . Thus,

$$P(h) = \sum_{f \in \mathcal{K}} \alpha_f P_f(h) \tag{39}$$

$$\geq \min_{f \in \mathcal{K}} P_f(h) \tag{40}$$

$$= \min_{f \in \mathcal{K}} \max_{P \in \mathcal{M}_f} P(h) \tag{41}$$

Hence,  $\underline{P}(h) = \min_{P \in \mathcal{M}'} P(h) \geq \min_{f \in \mathcal{K}} \max_{P \in \mathcal{M}_f} P(h)$ , which combined with the above reverse inequality gives the required equality.

The fact that extremal values are reached in extreme points easily implies (iii).  $\square$

The following corollary gives the maximal distance between two extensions of  $\underline{P}|_{\mathcal{K}}$ .

**Corollary 6.** *Let  $\underline{E}$  be the natural extension and  $\underline{P}$  and  $\underline{P}'$  two other extensions of  $\underline{P}|_{\mathcal{K}}$ , and  $h \in \mathcal{L}$  a gamble. Then*

$$|\underline{P}(h) - \underline{P}'(h)| \leq \min_{f \in \mathcal{K}} \max_{E \in \mathcal{E}_f} E(h) - \underline{E}(h). \quad (42)$$

*Proof.* The inequality is a direct consequence of Corollary 5(iii) and Eq. (37).  $\square$

An estimate of the upper bound to the maximal distance between  $\underline{E}$  and  $\underline{P}$  is now the following:

$$d(\underline{P}, \underline{E}) = \max_{\substack{h \in \mathcal{L} \\ \|h\|=1}} |\underline{P}(h) - \underline{E}(h)| \quad (43)$$

because of  $\underline{P}(h) \geq \underline{E}(h)$  :

$$= \max_{\substack{h \in \mathcal{L} \\ \|h\|=1}} \underline{P}(h) - \underline{E}(h) \quad (44)$$

using Corollary 5(iii):

$$\leq \max_{\substack{h \in \mathcal{L} \\ \|h\|=1}} \left\{ \min_{f \in \mathcal{K}} \max_{F \in \mathcal{E}_f} F(h) - \min_{E \in \mathcal{E}(\mathcal{M})} E(h) \right\} \quad (45)$$

$$= \max_{E \in \mathcal{E}(\mathcal{M})} \min_{f \in \mathcal{K}} \max_{F \in \mathcal{E}_f} \max_{\substack{h \in \mathcal{L} \\ \|h\|=1}} F(h) - E(h) \quad (46)$$

$$= \max_{E \in \mathcal{E}(\mathcal{M})} \min_{f \in \mathcal{K}} \max_{F \in \mathcal{E}_f} d(E, F) \quad (47)$$

Notice that given  $h$  the RHS of (46) is maximized if  $E(h) = \underline{E}(h)$ , or in other words, if  $h \in N_{\mathcal{M}}(E)$  (see (19)). Therefore:

$$d(\underline{P}, \underline{E}) \leq \max_{E \in \mathcal{E}(\mathcal{M})} \min_{f \in \mathcal{K}} \max_{F \in \mathcal{E}_f} \max_{h \in N_{\mathcal{M}}(E)} \frac{d(E(h), F(h))}{\|h\|} \quad (48)$$

$$= \max_{E \in \mathcal{E}(\mathcal{M})} \min_{f \in \mathcal{K}} \max_{F \in \mathcal{E}_f} d_E(E, F). \quad (49)$$

As follows from the above propositions, the inequalities above can become equalities with an appropriate  $\underline{P}$ .

The minimum of the RHS of the last equation can be found as a solution of the optimization problem (32)–(34).

To use equation (48) directly for calculating the maximal distance, one would need to examine every extreme point  $E$  of  $\mathcal{M}$  and every face of  $\mathcal{M}$  and then select the face whose most distant extreme point from  $E$  is nearest to  $E$ . This obviously requires a lot of redundant analysis, because most of the faces of  $\mathcal{M}$  can be eliminated by applying a much faster criterion.

**Proposition 4.** *Let  $E$ ,  $F$  and  $F'$  be linear previsions and  $h \in N_{\mathcal{M}}(E)$ . Suppose that  $F'(f_i) \geq F(f_i)$  for all the elements of the positive basis of  $N_{\mathcal{M}}(E)$ . Then  $F'(h) \geq F(h)$  for every  $h \in N_{\mathcal{M}}(E)$ .*

*Proof.* An easy consequence of the fact that every element  $h \in N_{\mathcal{M}}(E)$  is a positive combination of elements  $f_i$ , for  $i \in I$ .  $\square$

In the circumstances described by the above proposition we will say that an extreme point  $F'$  *dominates*  $F$  on  $N_{\mathcal{M}}(E)$ .

## 5 Algorithm

### 5.1 Outline

The algorithm for finding the maximal distance between a lower prevision and the natural extension of its restriction to  $\mathcal{K}$  is based on equation (48). As shown in previous sections, the maximal distance can be computed in terms of extreme points  $\mathcal{E}$  of  $\mathcal{M}$ . Efficient algorithms for finding the extreme points are known ([4, 6, 7, 9]), whose worst case complexity is estimated to  $O(n^2 dv)$  (see e.g. [3]), where  $n$  is the number of constraints,  $d$  the dimension, and  $v$  the number of extreme points (vertices).

For every extreme point  $E$  we need to find the face whose most distant point is nearest to  $E$ . It is reasonable to start with the faces nearest to  $E$ , which certainly are those, whose extreme points include  $E$ . Finding the maximal distances for those faces gives a reasonable estimate of the maximal distance for the given  $E$ ; however, there might exist faces whose most distant points are nearer than that. Even though there is no obvious way to select those faces, they must certainly not contain extreme points that dominate all the extreme points of any face, whose distances to  $E$  have already been calculated. Filtering out the set of faces containing dominating extreme points is relatively fast operation that significantly reduces the number of faces left to examine.

**Remark 3.** Non-negativity constraints of the form  $p \cdot 1_x \geq 0$  may appear among (13), where  $1_x$  does not belong to  $\mathcal{K}$ . This means that the corresponding credal set  $\mathcal{M}$  contains faces of the form  $\mathcal{M}_x = \{P \in \mathcal{M}: P \cdot 1_x = 0\}$ , which must be excluded from the set of faces that necessarily intersect the



credal set of  $\underline{P}$ , and whose distance from  $E$  is a candidate for the smallest distance. In the algorithm, this requires an additional test whether a face corresponds to an original gamble. However, to keep the presentation simple, we will sometimes assume that every face of  $\mathcal{M}$  is of the form  $\mathcal{M}_f$ , where  $f \in \mathcal{K}$ .

## 5.2 Parts of the algorithm

*Finding extreme points.* This step applies one of the existing algorithms for finding extreme points of a convex polyhedron. The inputs of the function `GENERATEEXTREMEPOINTS(gmb, lpr)` are the set of gambles `gmb`, which is in the form of an  $n \times s$  matrix, where  $n$  is the number of gambles and  $s$  is their dimension and `lpr` is the column vector of their lower previsions. The output is the set of extreme points in the form of a  $v \times s$  matrix `V`.

*Complete constraints.* This part adds the non-negativity constraints of the form  $p \cdot 1_x \geq 0$  and then removes the possible loose constraints. A constraint  $p \cdot f \geq \underline{P}(f)$  is loose if there is no extreme point  $E \in \mathcal{E}$  such that  $E(f) = \underline{P}(f)$ . Note that except for the non-negativity constraints, coherence of the lower prevision in principle prevents the existence of loose constraints. The function `REMOVEDUNDANTCONSTRAINTS(fn, lpr, EP)` returns `fn` and `lpr` inducing the same set of extreme points `EP` and without loose constraints. If any added non-negativity constraints remain non-loose, the corresponding gambles must be excluded from the set of faces whose maximal distance from extreme points is calculated.

*Finding the distance between the extreme points  $E$  and  $F$  (Algorithm 1).* Maximizes the expression (25) on the normal cone  $N_{\mathcal{M}}(E)$ . The problem translates to solving quadratic programming problem (32)–(34). The inputs are the extreme points `E` and `P` as vectors of length  $s$  and the gambles that form the basis of the normal cone of `E` in the form of  $I \times s$  matrix `fpos`.

---

**Algorithm 1** Function: normed distance

---

```

1: function NORMEDDISTANCE(E, P, fpos)
2:   D ← P-E
3:   nD ←  $\sqrt{D \cdot D}$  ▷ norm of D
4:   set Dmat to be the matrix compatible with fpos with all rows equal D
5:   λ ← fpos · Dmat/nD ▷ λ becomes a column vector
6:   u ← fpos - λ · Dmat ▷ u becomes a matrix
7:   Pi ← nD2λλt + uut
8:   dist ← min αtPiα subject to nD2λ · α = 1, α ≥ 0
9:   ▷ minimize over the set of s-dimensional vectors
10:  ▷ using quadratic programming
11:  return dist
12: end function

```

---

*Finding extreme points dominating/dominated-by an extreme point.* According to Proposition 4, an extreme point that dominates another extreme point on the basis of a normal cone, dominates it in the entire cone. This fact allows optimizing the algorithm, since several optimization steps are not needed in the case of dominance. The function is called `DOMINATEDEXTREMEPOINTS(E, points, fpos)`. The inputs are an extreme point `E`, a list of extreme points `points` and a list of gambles `fpos`. The output is a set of indices of those extreme points from the list `points` that are dominated by `E`.

*Filter faces.* Filters the faces containing extreme points that dominate entire faces already analysed. These faces can be left out of further analysis. Function `FILTERDOMINATINGFACES(fcs, domP)` returns those among faces `fcs`, whose set of extreme points does not intersect the set of dominating extreme points `domP`.

*Find the maximal distance (Algorithm 2).* Finds the maximum of (48). As inputs we take a set of constraints in the form of gambles `gmb` and their lower previsions `lpr`. The function returns the maximal possible distance between any two extensions of these assessments to the set of all gambles. In principle, our algorithm calculates the distances between all extreme points that lie in a common face  $\mathcal{M}_f$  for some  $f \in \mathcal{K}$ . After that it also checks all other faces whose none of the extreme points are dominated by at least one extreme point already tested. This part is meant to ensure that the calculated distance is exact, although, such faces rarely exist or even improve the calculated distances. If only a good estimate is needed, this step may as well be omitted. Still, even if it remains, it does not significantly add to the computation time, since it mostly uses fast matrix algebra routines.

Contrary, excluding dominated points (lines 26–28) significantly improves the efficiency of the algorithm. Empirical testing shows that only distances for a fraction of points then need to be calculated. Since the calculation of the distances is by far the slowest part of the algorithm, this significantly shortens the run-time of the algorithm.

### 5.3 Complexity estimation of the algorithm

Space complexity is determined by the number of extreme points, which depends on the shape of the credal set. For general lower previsions, even in low dimensions their number can be arbitrarily large; however, in the case of lower-upper probability pairs the upper bound for the number of extreme points is reported to be  $s!$ , where  $s$  is equal to the number of elements of the probability space (see [18]). Special classes of coherent lower-upper probability pairs have also been analysed with the focus to their extreme points in [11, 12]. The time complexity of the enumeration of extreme points of polyhedra is  $O(n^2dv)$  (see [3]), where  $n$  is the number of constraints,  $d$  the dimension, which is typically equal to  $s - 1$ , and  $v$  is the number of vertices.

---

**Algorithm 2** Function: find maximal distance
 

---

```

1: function MAXIMALDISTANCE(gmb, lpr)
2:    $V \leftarrow \text{GenerateExtremePoints}(\text{fn}, \text{lpr})$ 
3:    $\text{maxDist} = 0$ 
4:   for each  $E$  in  $V$  do
5:      $\text{minDist} \leftarrow \infty$ 
6:      $\text{fpos} \leftarrow \{f \in \text{gmb} : E \cdot f = \text{lpr}(f)\}$ 
7:        $\triangleright$  get support gambles of faces whose extreme point is  $E$ 
8:        $\triangleright$  they constitute positive basis of the normal cone
9:     for each  $f \in \text{fpos} \setminus \text{added non-negativity constraints}$  do
10:       $V_f \leftarrow \{P \in V : E \cdot f = \text{lpr}(f)\}$ 
11:         $\triangleright$  get all extreme points of the face  $\mathcal{M}_f$ 
12:       $V_f \leftarrow \text{SortByEuclidianDistanceToE}(V_f)$ 
13:         $\triangleright$  we start with the point that is most distant from  $E$ 
14:       $\text{maxFaceDist} \leftarrow 0$ 
15:         $\triangleright$  maximal distance to an extreme point of the current face
16:       $\text{dominated} \leftarrow \emptyset$ 
17:      for each  $P \in V_f$  do
18:        if  $P \in \text{dominated}$  then
19:           $d \leftarrow \text{maxFaceDist}$ 
20:             $\triangleright$  distance calculation on
21:             $\triangleright$  dominated faces is unnecessary
22:        else
23:           $d \leftarrow \text{NormedDistance}(E, P, \text{fpos})$ 
24:        end if
25:         $\text{maxFaceDist} \leftarrow \max(\text{maxFaceDist}, d)$ 
26:         $V_fD \leftarrow \text{DominatedExtremePoints}(P, V_f, \text{fpos})$ 
27:         $\text{dominated} \leftarrow \text{dominated} \cup V_fD$ 
28:         $\triangleright$  exclude all points dominated by  $P$  on the normal cone
29:      end for
30:       $\text{minDist} \leftarrow \min(\text{minDist}, \text{maxFaceDist})$ 
31:       $\text{filtGmb} \leftarrow \text{FilterDominatedFaces}(\text{gmb}, \text{tested})$ 
32:         $\triangleright$  filter gambles dominating already tested points
33:      repeat steps 10–30 with  $\text{filtGmb}$  in place of  $V_f$ 
34:    end for
35:     $\text{maxDist} \leftarrow \max(\text{maxDist}, \text{minDist})$ 
36:  end for
37:  return  $\text{maxDist}$ 
38: end function

```

---

A more severe obstacle than space complexity is its time complexity. By far the slowest part of our algorithm is the calculation of the distance between extreme points (Algorithm 2, line 23) described in Algorithm 1. Our complexity analysis will therefore focus on the number of calls to this routine. The time complexity of the routine is around  $o(s^3)$ , but since  $s$  will typically be small compared to other variables, we may regard it as constant, and do the time complexity analysis based on other factors. The most important factor is certainly the number of extreme points, which we will denote by  $v$ ; the number of constraints, which roughly coincides with the number of facets, will be denoted by  $n$  and the dimension by  $d$ , typically  $d = s - 1$ .

Typically an extreme point is a solution of a system of  $d$  equations forming constraints. This means that every extreme point is most usually adjacent to  $d$  facets. The number of vertices per facet is then on average equal to  $\frac{vd}{n}$ , and since the distance to all vertices of faces adjacent to a vertex must be calculated, this gives  $\frac{vd^2}{n}$  vertices. Vertices that were counted twice, because they lie on two facets, must be subtracted from this number. The distance must be calculated for every vertex, so this number must be multiplied by  $v$ . Depending on the ratio  $\frac{v}{n}$ , the number of pairs for which the quadratic programming routine `NormedDistance` must be called is bounded above by  $d^2v^2/n$ . In practice, the number of calls is significantly reduced by eliminating dominated extreme points.

The time complexity is therefore exponential as a function of  $s$  and polynomial as a function of  $v$ . Practical testing shows that even for relatively low dimensional cases the algorithm's complexity is high. The complexity increases with the number of constraints, which is the number of gambles in  $\mathcal{K}$ . Notice however that with the size of  $|\mathcal{K}|$  the accuracy of the approximation of the partially specified lower prevision increases. Therefore, the maximal possible distance, which is the maximal possible error of the approximation, is more important in the cases where the number of estimates is low; and in those cases the computational complexity of the algorithm is lower.

## 5.4 Numerical testing

For numerical testing we implemented the algorithm on a sample of 10 randomly generated lower probabilities on probability spaces of sizes  $s = |\mathcal{X}| = 3, 4, 5, 6$  and 7. The constraints are formed by the lower probabilities of non-trivial subsets, whose number is  $2^s - 2$ ; the number of extreme points was in general close to the maximal possible number, which is  $s!$ . According to the complexity estimation from the previous section the upper bound for the number of calls without filtering the dominated extreme points would

$ \mathcal{X} $	ext. pts	dist. calculated	dist. needed	ratio
3	5,9	11,8	11,8	100,00%
4	23,6	124,2	300,4	41,34%
5	101,2	1697,2	6249,4	27,16%
6	592,3	31179,7	187453,2	16,63%
7	2744,7	586728,0	3911809,6	15,00%

Table 1: Test results by the sample space size: average number of extreme points; calls to quadratic programming routine; adjacent extreme points; ratio between the number of calls and the number of adjacent extreme points.

be

$$\frac{(s-1)^2(s!)^2}{2^s-2} \quad (50)$$

plus the number of extreme points of non-dominated non-adjacent faces  $\mathcal{M}_f$ . The algorithm was designed to count the actual number of distances between extreme points that need to be considered (Table 1, column 4), which slightly differs from the estimated value above. Further, Table 1 displays the average number of extreme points, which is also only slightly smaller than  $s!$ ; the number of calls to the quadratic programming routine to calculate distances between pairs of extreme points; the number of distances that had to be considered (not all of them need actually be considered because the dominated ones can be quickly eliminated from the analysis); and the percentage of needed distances that actually had to be calculated using the quadratic programming routine. As expected, this percentage drops with the increase of dimension.

## 6 Conclusions

The results obtained in this paper give a strict bound to the maximal error of an approximation of a coherent lower prevision with its restriction to a set of gambles. This subsumes approximations of coherent lower previsions with more specific models, such as coherent lower probabilities. Such approximations are very common in the applications of imprecise probabilities, and our results give a first attempt to evaluate their errors.

We have also provided an algorithm which calculates the bound based on calculating the normed distances between extreme points of a credal set. Since the number of extreme points grows rapidly with the dimension of the probability space, the computational complexity of the algorithm is in general high.

The high computational complexity of the algorithm presents an obstacle that might hinder its practical applicability. Therefore, a direction of

further research is finding more efficient algorithms based on the insights from the theoretical part of the paper. Another related question that could be examined with the help of the concept developed in this paper is how to choose the most optimal set of gambles to approximate an unknown coherent lower prevision with minimal error.

## References

- [1] Antonucci, Alessandro, & Cuzzolin, Fabio. 2010. Credal sets approximation by lower probabilities: application to credal networks. *Pages 716–725 of: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer.
- [2] Augustin, Thomas, Coolen, Frank PA, de Cooman, Gert, & Troffaes, Matthias CM. 2014. *Introduction to imprecise probabilities*. John Wiley & Sons.
- [3] Avis, David, & Fukuda, Komei. 1992. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry*, **8**(3), 295–313.
- [4] Chen, Pey-Chun, Hansen, Pierre, & Jaumard, Brigitte. 1991. On-line and off-line vertex enumeration by adjacency lists. *Operations Research Letters*, **10**(7), 403–409.
- [5] de Cooman, Gert, Hermans, Filip, & Quaeghebeur, Erik. 2009. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, **23**(4), 597–635.
- [6] Dyer, Martin E, & Proll, Les G. 1977. An algorithm for determining all extreme points of a convex polytope. *Mathematical Programming*, **12**(1), 81–96.
- [7] Dyer, M.E., & Proll, L.G. 1982. An improved vertex enumeration algorithm. *European Journal of Operational Research*, **9**(4), 359 – 368.
- [8] Gruber, Peter. 2007. *Convex and Discrete Geometry*. Springer-Verlag Berlin Heidelberg.
- [9] Matheiss, TH, & Rubin, David S. 1980. A survey and comparison of methods for finding all vertices of convex polyhedral sets. *Mathematics of operations research*, **5**(2), 167–185.
- [10] Miranda, Enrique. 2008. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, **48**(2), 628 – 658. In Memory of Philippe Smets (1938-2005).

- [11] Miranda, Enrique, & Destercke, Sbastien. 2015. Extreme points of the credal sets generated by comparative probabilities. *Journal of Mathematical Psychology*, **64-65**, 44 – 57.
- [12] Miranda, Enrique, Couso, Inés, & Gil, Pedro. 2003. Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning*, **33**(1), 95–115.
- [13] Quaeghebeur, Erik. 2015. Characterizing coherence, correcting incoherence. *International Journal of Approximate Reasoning*, **56, Part B**, 208 – 223. Eighth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2013).
- [14] Troffaes, Matthias CM, & De Cooman, Gert. 2014. *Lower previsions*. John Wiley & Sons.
- [15] Škulj, Damjan. 2009. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, **50**(8), 1314–1329.
- [16] Škulj, Damjan. 2015. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, **250**(0), 165 – 180.
- [17] Škulj, Damjan, & Hable, Robert. 2013. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, **76**(1), 107–133.
- [18] Wallner, Anton. 2007. Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning*, **44**(3), 339 – 357.